MAXIMUM LIKELIHOOD ESTIMATION FOR VECTOR AUTOREGRESSIVE MOVING AVERAGE MODELS

BY

T. W. ANDERSON

TECHNICAL REPORT NO. 35
JULY 1978

PREPARED UNDER CONTRACT N00014-75-C-0442
(NR-042-034)
OFFICE OF NAVAL RESEARCH

THEODORE W. ANDERSON, PROJECT DIRECTOR

DEPARTMENT OF STATISTICS
STANFORD UNIVERSITY
STANFORD, CALIFORNIA

MAXIMUM LIKELIHOOD ESTIMATION FOR VECTOR AUTOREGRESSIVE MOVING AVERAGE MODELS

BY

T. W. ANDERSON

Technical Report No. 35

July 1978

PREPARED UNDER CONTRACT N00014-75-C-0442

(NR-042-034)

OFFICE OF NAVAL RESEARCH

Theodore W. Anderson, Project Director

DEPARTMENT OF STATISTICS

STANFORD UNIVERSITY

STANFORD, CALIFORNIA

Maximum Likelihood Estimation for Vector Autoregressive Moving Average Models

T. W. Anderson
Stanford University

ABSTRACT

The vector autoregressive moving average model is a multivariate stationary stochastic process $\{\underset{\sim}{y}_t\}$ satisfying

$$\sum_{k=0}^{p} \underset{\sim}{B}_k \underset{\sim}{y}_{t-k} = \sum_{g=0}^{q} \underset{\sim}{A}_g \underset{\sim}{v}_{t-g} ,$$

where the unobservable multivariate process $\{\underset{\sim}{v}_t\}$ consists of independently identically distributed random vectors. The coefficient matrices and the covariance matrix of $\underset{\sim}{v}_t$ are to be estimated from an observed sequence $\underset{\sim}{y}_1, \ldots, \underset{\sim}{y}_T$. Under the assumption of normality the method of maximum likelihood is applied to likelihoods suitably modified for techniques in the frequency and time domains. Newton-Raphson and scoring iterative methods are presented.

KEY WORDS: Maximum likelihood estimation, vector autoregressive moving average, Newton-Raphson, scoring, information matrix, time series analysis.

Maximum Likelihood Estimation for Vector

Autoregressive Moving Average Models

T. W. Anderson*

Stanford University

1. Introduction. The purpose of this paper is to review and relate several methods of estimating the coefficients of a vector-valued autoregressive moving average process. These procedures are based on the application of the Newton-Raphson method or the scoring method to modifications of the likelihood function of a Gaussian model. To some extent this paper does for the multivariate process what Anderson (1977) did for the univariate process.

The observable m-component vector-valued autoregressive moving average process $\{y_t\}$ satisfies

$$\sum_{k=0}^{p} B_k y_{t-k} = \sum_{g=0}^{q} A_g v_{t-g} , \tag{1.1}$$

$t = \ldots, -1, 0, 1, \ldots$, where the sequence $\{v_t\}$ consists of unobservable independently identically distributed random m-component vectors with $\mathcal{E} v_t = 0$ and $\mathcal{E} v_t v_t' = \Sigma$, assumed nonsingular and $B_0, \ldots, B_p$ and $A_0, \ldots, A_q$ are $m \times m$ matrices. To avoid indeterminacy we require $B_0 = A_0 = I_m$. We take $\mathcal{E} y_t = 0$ (with no loss of generality if $\mathcal{E} y_t$ is known) because we are interested in the covariance structure as it depends on the coefficient matrices.

*I am indebted to Fereydoon Ahrabi and Paul Shaman for assistance and advice on preparing this paper.

Let

$$(1.2) \qquad \underset{\sim}{B}(z) = \sum_{k=0}^{p} \underset{\sim}{B}_k z^k , \quad \underset{\sim}{A}(z) = \sum_{g=0}^{q} \underset{\sim}{A}_g z^g .$$

We shall assume that the zeros of $|\underset{\sim}{B}(z)|$ and $|\underset{\sim}{A}(z)|$ are greater than 1 in absolute value.

For an arbitrary stationary m-component process $\{\underset{\sim}{y}_t\}$ with $\mathcal{E}\underset{\sim}{y}_t = \underset{\sim}{0}$ we can define the covariance sequence

$$(1.3) \qquad \underset{\sim}{\Sigma}_h = \mathcal{E}\underset{\sim}{y}_t \underset{\sim}{y}'_{t+h} ,$$

$h = \ldots, -1, 0, 1, \ldots$, with $\underset{\sim}{\Sigma}_{-h} = \underset{\sim}{\Sigma}'_h$. If the series converges, the spectral density is

$$(1.4) \qquad \underset{\sim}{f}(\lambda) = \frac{1}{2\pi} \sum_{h=-\infty}^{\infty} e^{-i\lambda h} \underset{\sim}{\Sigma}_h ,$$

with $\underset{\sim}{f}'(\lambda) = \bar{\underset{\sim}{f}}(\lambda)$, where the bar denotes complex conjugate; that is, $\underset{\sim}{f}(\lambda)$ is Hermitian. Then

$$(1.5) \qquad \underset{\sim}{\Sigma}_k = \int_{-\pi}^{\pi} e^{i\lambda k} \underset{\sim}{f}(\lambda)\, d\lambda ,$$

where the right-hand side consists of the matrix with each element being the integral of the corresponding element of $e^{i\lambda k} \underset{\sim}{f}(\lambda)$. For the autoregressive moving average process defined by (1.1) the matrix-valued spectral density is

$$(1.6) \qquad \underset{\sim}{f}(\lambda) = \frac{1}{2\pi} \underset{\sim}{B}(e^{i\lambda})^{-1} \underset{\sim}{A}(e^{i\lambda}) \underset{\sim}{\Sigma} \underset{\sim}{A}^*(e^{i\lambda}) \underset{\sim}{B}^*(e^{i\lambda})^{-1} ,$$

where * denotes complex conjugate transpose. (See Anderson (1971) and Hannan (1970) for general discussion of scalar and vector processes, respectively.)

Any stationary process $\{y_t\}$ with finite second-order moments determines the covariance sequence $\{\Sigma_h\}$ which in turn determines the spectral density when it exists and (1.4) converges in a suitable sense. Conversely, if the process is Gaussian either the covariance sequence or the spectral density describes the process. In particular, an autoregressive moving average process satisfying (1.1) with $\mathcal{E} v_t v_t' = \Sigma$ determines $f(\lambda)$ by (1.6). Does that $f(\lambda)$ uniquely determine the matrix Σ and the matrix polynomials $B(z)$ and $A(z)$? The answer is No without further conditions. What are uniquely determined are the matrix Σ and the rational transfer function, which can be written $B(z)^{-1} A(z)$. However, we obtain the same transfer function from $\tilde{B}(z)^{-1} \tilde{A}(z)$, where $\tilde{A}(z)$ and $\tilde{B}(z)$ are defined by

$$\tilde{A}(z) = C(z) A(z), \quad \tilde{B}(z) = C(z) B(z), \tag{1.7}$$

where $C(z)$ is another matrix of polynomials which is nonsingular for $|z| = 1$. The matrix $C(z)$ is called a common left divisor of $\tilde{A}(z)$ and $\tilde{B}(z)$, and $\tilde{A}(z)$ and $\tilde{B}(z)$ are right multiples of $C(z)$. The greatest common left divisor of two polynomial matrices is a common left divisor and any other common left divisor has this matrix as a right multiple. The greatest common left divisor is not unique; it can be multiplied on the right by a unimodular matrix (that is, a polynomial matrix with constant determinant). We can ask that a greatest common left divisor be I_m, but to eliminate the indeterminacy of multiplication by a unimodular matrix another condition should be added. One such condition is that the rank of $(B_p\ A_q)$ is m [Hannan (1969a)]. Other conditions can replace this last one (which is not a necessary condition); see Hannan (1971) and Kashyap and Nasburg (1974).

The statistical inference problem is to estimate $\underset{\sim}{B}_1, \ldots, \underset{\sim}{B}_p$, $\underset{\sim}{A}_1, \ldots, \underset{\sim}{A}_q$, and $\underset{\sim}{\Sigma}$ (p and q given) on the basis of T observations $\underset{\sim}{y}_1, \ldots, \underset{\sim}{y}_T$. The method of maximum likelihood can be considered under the assumption that the process is Gaussian. The problem, which is the optimization of a complicated objective function, can be solved numerically. However, we consider modifying the model so that the Newton-Raphson or scoring method can be used in a straight-forward way. In the time domain we modify the likelihood function by treating the variables before the period of observation as zero; that is, $\underset{\sim}{y}_0 = \cdots = \underset{\sim}{y}_{1-p} = \underset{\sim}{0}$ and $\underset{\sim}{v}_0 = \cdots = \underset{\sim}{v}_{1-q} = \underset{\sim}{0}$. This case is a little simpler than that of Reinsel (1976), who treats $\underset{\sim}{y}_1, \ldots, \underset{\sim}{y}_p$ as fixed and $\underset{\sim}{v}_{p+1-q} = \cdots = \underset{\sim}{v}_p = \underset{\sim}{0}$. Tunnicliffe Wilson (1973) proposed an equivalent procedure without specifying the choice of initial values of the variables.

In the frequency domain we use the sample spectral density (periodogram)

$$(1.8) \qquad \underset{\sim}{I}(\lambda) = \frac{1}{2\pi T} \sum_{t,s=1}^{T} \underset{\sim}{y}_t \underset{\sim}{y}_s' e^{i\lambda(t-s)}$$

$$= \frac{1}{2\pi} \sum_{h=-(T-1)}^{T-1} e^{-i\lambda h} \underset{\sim}{c}_h ,$$

where

$$(1.9) \qquad \underset{\sim}{c}_h = \frac{1}{T} \sum_{t=1}^{T-h} \underset{\sim}{y}_t \underset{\sim}{y}_{t+h}' , \quad h = 0, 1, \ldots, T-h ,$$

$$= \underset{\sim}{c}_{-h}'$$

Dunsmuir and Hannan (1976) have shown that the logarithm of the likelihood function for the observations on the Gaussian process can be approximated by a constant plus

$$(1.10) \qquad -\frac{T}{2}\log|\Sigma| \quad -\frac{1}{2}\sum_{t=1}^{T} \operatorname{tr} f^{-1}(\lambda_t)\, I(\lambda_t) \ ,$$

where $\lambda_t = (2\pi/T)t$, $t = 1, \ldots, T$, and $f(\lambda)$ is given by (1.6) . Hannan (1969b) used this approach for the strictly moving average process, and Akaike (1973) and Nicholls (1976) followed this approach for the autoregressive moving average model (with exogenous variables included). This logarithm of the likelihood function can also be obtained by modifying the model (1.1) by setting $y_0 = y_T, \ldots, y_{1-p} = y_{T-p+1}$ and $v_0 = v_T, \ldots, v_{1-q} = v_{T-q+1}$ [Anderson (1977)].

In general, if the likelihood of θ based on the observation of x is $L(x,\theta)$, then the Taylor's series expansion of $\log L(x,\theta)$ yields the equation

$$(1.11) \qquad -\left.\frac{\partial^2 \log L(x,\theta)}{\partial\theta\ \partial\theta'}\right|_{\theta=\hat{\theta}_0} (\hat{\theta}_1-\hat{\theta}_0) = \left.\frac{\partial}{\partial\theta}\log L(x|\theta)\right|_{\theta=\hat{\theta}_0} ,$$

which is to be solved for $\hat{\theta}_1$ given an initial estimate $\hat{\theta}_0$; this is the Newton-Raphson method. If $\hat{\theta}_0$ is consistent of probability order $1/\sqrt{T}$, then in most models, $\hat{\theta}_1$ is consistent, asymptotically normal, and asymptotically efficient. It is customary to iterate (1.11), using a solution $\hat{\theta}_1$ as the initial estimate $\hat{\theta}_0$ in the next step.

The information matrix is

$$(1.12) \qquad -\left[\mathcal{E}_\theta \frac{\partial^2 \log L(X|\theta)}{\partial\theta\ \partial\theta'}\right]_{\theta=\hat{\theta}_0} ,$$

where X is the random vector on which x is the observation. The method of scoring consists of the estimation procedure (1.11) with the information matrix replacing the matrix of second partial derivatives. Iteration is usually carried out.

To write linear equations for the elements of matrices it is convenient to use the "vec" of a matrix.

Definition. If $\underset{\sim}{C} = (\underset{\sim}{c}_1, \ldots, \underset{\sim}{c}_n)$,

(1.13) $$\text{vec } \underset{\sim}{C} = \begin{pmatrix} \underset{\sim}{c}_1 \\ \vdots \\ \underset{\sim}{c}_n \end{pmatrix} .$$

We use the following result:

(1.14) $$\text{vec } \underset{\sim}{A}\underset{\sim}{B}\underset{\sim}{C} = (\underset{\sim}{C}' \otimes \underset{\sim}{A}) \text{ vec } \underset{\sim}{B} ,$$

which is easily verified by writing out the two sides; here $\otimes$ denotes the Kronecker product. (See, for example, Marcus and Minc (1944), p. 9.) Accordingly, we let

(1.15) $$\text{vec } (\underset{\sim}{A}_1 \ldots \underset{\sim}{A}_q) = \underset{\sim}{\alpha} , \quad \text{vec } (\underset{\sim}{B}_1 \ldots \underset{\sim}{B}_p) = \underset{\sim}{\beta} .$$

2. Estimation in the frequency domain. The equations for estimating $\underset{\sim}{\alpha}$ and $\underset{\sim}{\beta}$ given initial estimates of $\underset{\sim}{\alpha}$, $\underset{\sim}{\beta}$, and $\underset{\sim}{\Sigma}$ are

(2.1) $$\begin{bmatrix} \hat{\underset{\sim}{\Phi}}_0 & \hat{\underset{\sim}{\Omega}}_0 \\ \hat{\underset{\sim}{\Omega}}'_0 & \hat{\underset{\sim}{\Psi}}_0 \end{bmatrix} \begin{bmatrix} \hat{\underset{\sim}{\alpha}}_1 - \hat{\underset{\sim}{\alpha}}_0 \\ \hat{\underset{\sim}{\beta}}_1 - \hat{\underset{\sim}{\beta}}_0 \end{bmatrix} = \begin{bmatrix} \hat{\underset{\sim}{q}}_0 \\ \hat{\underset{\sim}{p}}_0 \end{bmatrix} .$$

The matrices $(1/T)\hat{\underset{\sim}{\Phi}}_0$, $(1/T)\hat{\underset{\sim}{\Omega}}_0$, and $(1/T)\hat{\underset{\sim}{\Psi}}_0$ are initial estimates of the corresponding submatrices of the (limiting) average information matrix

(2.2)
$$\underset{\sim}{\Phi} = \begin{bmatrix} \underset{\sim}{\Phi}_{11} & \cdots & \underset{\sim}{\Phi}_{1q} \\ \vdots & & \vdots \\ \underset{\sim}{\Phi}_{q1} & \cdots & \underset{\sim}{\Phi}_{qq} \end{bmatrix} ,$$

(2.3)
$$\underset{\sim}{\Omega} = \begin{bmatrix} \underset{\sim}{\Omega}_{11} & \cdots & \underset{\sim}{\Omega}_{1p} \\ \vdots & & \vdots \\ \underset{\sim}{\Omega}_{q1} & \cdots & \underset{\sim}{\Omega}_{qp} \end{bmatrix} ,$$

(2.4)
$$\underset{\sim}{\Psi} = \begin{bmatrix} \underset{\sim}{\Psi}_{11} & \cdots & \underset{\sim}{\Psi}_{1p} \\ \vdots & & \vdots \\ \underset{\sim}{\Psi}_{p1} & \cdots & \underset{\sim}{\Psi}_{pp} \end{bmatrix} ,$$

where the individual submatrices are square of order m^2, and $\underset{\sim}{\hat{q}}_0$ and $\underset{\sim}{\hat{p}}_0$ are composed of the first derivatives of the logarithm of the likelihood function evaluated at the initial estimates.

Let

(2.5)
$$\underset{\sim}{f}^u(\lambda) = \frac{1}{2\pi} \underset{\sim}{A}(e^{i\lambda}) \underset{\sim}{\Sigma} \underset{\sim}{A}^*(e^{i\lambda}) ,$$

which is the spectral density of the moving average part of (1.1) and is derived from the right-hand side. Then we can write the submatrices of the (limiting) average information matrix as

(2.6)
$$\underset{\sim}{\Phi}_{gh} = \frac{1}{(2\pi)^2} \int_{-\pi}^{\pi} \left[\underset{\sim}{\Sigma} \otimes \underset{\sim}{f}^{u'}(\lambda)^{-1} \right] e^{i(g-h)\lambda} \, d\lambda ,$$

$$(2.7) \qquad \underset{\sim}{\Omega}_{g\ell} = -\frac{1}{(2\pi)^2}\int_{-\pi}^{\pi}\left[\underset{\sim}{\Sigma}\ \underset{\sim}{A}^*\ (e^{i\lambda})\underset{\sim}{B}^*\ (e^{i\lambda})^{-1} \otimes \underset{\sim}{f}^{u\prime}(\lambda)^{-1}\right] e^{i(g-\ell)\lambda}d\lambda \ ,$$

$$(2.8) \qquad \underset{\sim}{\Psi}_{k\ell} = \frac{1}{(2\pi)^2}\int_{-\pi}^{\pi}\left[\underset{\sim}{B}(e^{i\lambda})^{-1}\underset{\sim}{A}(e^{i\lambda})\ \underset{\sim}{\Sigma}\ \underset{\sim}{A}^*\ (e^{i\lambda})\ \underset{\sim}{B}^*\ (e^{i\lambda})^{-1} \otimes \underset{\sim}{f}^{u\prime}(\lambda)^{-1}\right] e^{i(k-\ell)\lambda}d\lambda \ ,$$

$$= \frac{1}{2\pi}\int_{-\pi}^{\pi}\left[\underset{\sim}{f}\ (\lambda) \ \otimes\ \underset{\sim}{f}^{u\prime}(\lambda)^{-1}\right] e^{i(k-\ell)\lambda}d\lambda \ .$$

Each submatrix depends on its indices only through the difference of the indices; that is, the matrices $\underset{\sim}{\Phi}$, $\underset{\sim}{\Omega}$, and $\underset{\sim}{\Psi}$ are block Toeplitz. Note that $\underset{\sim}{\Phi}$, the moving average part of the average information matrix, depends only on the moving average part of the process. (We emphasize that $\hat{\underset{\sim}{\Phi}}_0$ estimates $T\ \underset{\sim}{\Phi}$, etc.)

The submatrix of the full (limiting) average information matrix that involves $\underset{\sim}{\alpha}$ and $\underset{\sim}{\Sigma}$ and the submatrix that involves $\underset{\sim}{\beta}$ and $\underset{\sim}{\Sigma}$ are composed of zeros. The covariance matrix is estimated by a separate equation.

<u>Newton-Raphson</u> (Akaike, Hannan, and Nicholls). In the Newton-Raphson method in the frequency domain the matrices on the left-hand side of (2.1) are the second partial derivatives of the logarithm of the likelihood function with asymptotically negligible terms omitted; they can be obtained from (1.10) . Let

$$(2.9) \qquad \hat{\underset{\sim}{f}}{}^{u}_{0}\ (\lambda_t) = \frac{1}{2\pi}\ \hat{\underset{\sim}{A}}_0\ (e^{i\lambda_t})\ \hat{\underset{\sim}{\Sigma}}_0\ \hat{\underset{\sim}{A}}{}^*_0\ (e^{i\lambda_t})\ .$$

In $\hat{A}_0(e^{i\lambda})$ and $\hat{B}_0(e^{i\lambda})$ we use the initial estimates of the coefficient matrices. Then the estimates of $T\Phi_{gh}$, $T\Omega_{g\ell}$, and $T\Psi_{k\ell}$ are

$$(2.10)\quad \hat{\Phi}^{(0)}_{gh} = \sum_{t=1}^{T} \left[\hat{A}_0(e^{i\lambda_t})^{-1}\hat{B}_0(e^{i\lambda_t}) I(\lambda_t) \hat{B}^*_0(e^{i\lambda_t}) \hat{A}^*_0(e^{i\lambda_t})^{-1} \otimes \hat{f}^{u\prime}_0(\lambda_t)^{-1} \right] e^{i(g-h)\lambda_t},$$

$$(2.11)\quad \hat{\Omega}^{(0)}_{g\ell} = -\sum_{t=1}^{T} \left[\hat{A}_0(e^{i\lambda_t})^{-1}\hat{B}_0(e^{i\lambda_t}) I(\lambda_t) \otimes \hat{f}^{u\prime}_0(\lambda_t)^{-1} \right] e^{i(g-\ell)\lambda_t},$$

$$(2.12)\quad \hat{\Psi}^{(0)}_{k\ell} = \sum_{t=1}^{T} \left[I(\lambda_t) \otimes \hat{f}^{u\prime}_0(\lambda_t)^{-1} \right] e^{i(k-\ell)\lambda_t}.$$

The matrices $\hat{\Phi}_0$, $\hat{\Omega}_0$, and $\hat{\Psi}_0$ are again block Toeplitz. Each submatrix is made up of estimates of the matrices appearing in the (limiting) average information matrix.

The right-hand side of (2.1) consists of the partial derivatives of the logarithm of the (approximate) likelihood function with respect to the elements of α and β arranged in the forms of (column) vectors. The g-th subvector of $\hat{q}_0$ and the k-th subvector of $\hat{p}_0$ are

$$(2.13)\quad \hat{q}^{(0)}_g = \text{vec}\left[\sum_{t=1}^{T} \hat{f}^{u\prime}_0(\lambda_t)^{-1} \bar{\hat{B}}_0(e^{i\lambda_t}) I'(\lambda_t) \hat{B}'_0(e^{i\lambda_t}) A'_0(e^{i\lambda_t})^{-1} e^{i\lambda_t g} \right],$$

$$(2.14) \qquad \hat{\underset{\sim}{p}}_k^{(0)} = -\operatorname{vec}\left[\sum_{t=1}^{T} \hat{\underset{\sim}{f}}{}_0^{u'}(\lambda_t)^{-1} \bar{\hat{\underset{\sim}{B}}}_0(e^{i\lambda_t}) \underset{\sim}{I}'(\lambda_t) e^{-i\lambda_t k}\right].$$

Since the first partial derivatives of the logarithm of the likelihood function with respect to the elements of $\underset{\sim}{\beta}$ are linear in the elements of $\underset{\sim}{\beta}$, they can be set equal to zero. The solutions in terms of the initial estimates of $\underset{\sim}{\alpha}$ and $\underset{\sim}{\Sigma}$ constitute an alternative "initial estimate" of $\underset{\sim}{\beta}$ and can be used in (2.1). Then $\hat{\underset{\sim}{p}}_0 = \underset{\sim}{0}$ and the solution for $\hat{\underset{\sim}{\alpha}}_1$ is easier. This is what Hannan (1969b) and (1971) did in the scalar case and Nicholls did (1976) in the vector case.

<u>Scoring</u>. In the expressions for $T\underset{\sim}{\Phi}_{gh}$, $T\underset{\sim}{\Omega}_{g\ell}$, and $T\underset{\sim}{\Psi}_{k\ell}$, we replace the parameters by their initial estimates, multiply by 2π, and sum over t instead of integrate over λ. The resulting expressions are

$$(2.15) \qquad \hat{\underset{\sim}{\Phi}}{}_{gh}^{(0)} = \frac{1}{2\pi} \sum_{t=1}^{T} \left[\hat{\underset{\sim}{\Sigma}}_0 \otimes \hat{\underset{\sim}{f}}{}_0^{u'}(\lambda_t)^{-1}\right] e^{i(g-h)\lambda_t},$$

$$(2.16) \qquad \hat{\underset{\sim}{\Omega}}{}_{g\ell}^{(0)} = -\frac{1}{2\pi} \sum_{t=1}^{T} \left[\hat{\underset{\sim}{\Sigma}}_0 \hat{\underset{\sim}{A}}{}_0^*(e^{i\lambda_t}) \hat{\underset{\sim}{B}}{}_0^*(e^{i\lambda_t})^{-1} \otimes \hat{\underset{\sim}{f}}{}_0^{u'}(\lambda_t)^{-1}\right] e^{i(g-\ell)\lambda_t},$$

$$(2.17) \qquad \hat{\underset{\sim}{\Psi}}{}_{k\ell}^{(0)} = \frac{1}{2\pi} \sum_{t=1}^{T} \left[\hat{\underset{\sim}{B}}_0(e^{i\lambda_t})^{-1} \hat{\underset{\sim}{A}}_0(e^{i\lambda_t}) \hat{\underset{\sim}{\Sigma}}_0 \hat{\underset{\sim}{A}}{}_0^*(e^{i\lambda_t}) \hat{\underset{\sim}{B}}{}_0^*(e^{i\lambda_t})^{-1} \otimes \hat{\underset{\sim}{f}}{}_0^{u'}(\lambda_t)^{-1}\right] e^{i(k-\ell)\lambda_t}.$$

The right-hand sides of (2.1) are the same as for the Newton-Raphson method. The estimate for the covariance matrix is given by

$$(2.18) \qquad \hat{\Sigma}_1 = \frac{2\pi}{T} \sum_{t=1}^{T} \hat{A}_0(e^{i\lambda_t})^{-1} \hat{B}_0(e^{i\lambda_t}) I(\lambda_t) \hat{B}_0^*(e^{i\lambda_t}) \hat{A}_0^*(e^{i\lambda_t})^{-1} .$$

As $T \to \infty$, $\sqrt{T}\,[(\hat{\alpha}_1 - \alpha)', (\hat{\beta}_1 - \beta)']'$ has a limiting normal distribution with mean zero and covariance matrix

$$(2.19) \qquad \begin{bmatrix} \Phi & \Omega \\ \Omega' & \Psi \end{bmatrix}^{-1} .$$

Nicholls (1976) has given this result for the v_t's independently identically distributed and the estimates based on his modification of the Newton-Raphson method with consistent initial estimates of order $1/\sqrt{T}$ in probability. The result also holds for the scoring method. Dunsmuir and Hannan (1976) have justified their results under very general conditions on the v_t's. Nicholls (1977) showed that the estimates he derived were a kind of "three-stage realization" of the Newton-Raphson method.

3. Estimation in the time domain. It will be convenient to define

$$(3.1) \qquad Y = \begin{bmatrix} y_1' \\ \vdots \\ y_T' \end{bmatrix}, \quad V = \begin{bmatrix} v_1' \\ \vdots \\ v_T' \end{bmatrix},$$

which are $T \times m$ matrices, and

$$(3.2) \qquad \operatorname{vec} Y = y , \quad \operatorname{vec} V = v ,$$

which are $Tm \times 1$ vectors. When $y_0 = \dots = y_{1-p} = 0$ and $v_0 = \dots = v_{1-q} = 0$, the model can be written as

$$(3.3) \qquad \sum_{k=0}^{p} L^k Y B_k' = \sum_{g=0}^{q} L^g V A_g' ,$$

where

(3.4) $$\underset{\sim}{L} = \begin{bmatrix} \underset{\sim}{0} & 0 \\ & \\ \underset{\sim}{I}_{T-1} & \underset{\sim}{0} \end{bmatrix}$$

with $\underset{\sim}{L}^0 = \underset{\sim}{I}_T$. Since

(3.5) $$\text{vec } \underset{\sim}{L}^k \underset{\sim}{Y} \underset{\sim}{B}_k' = (\underset{\sim}{B}_k \otimes \underset{\sim}{L}^k) \text{ vec } \underset{\sim}{Y} ,$$

the modified model can be written

(3.6) $$\underset{\sim}{\mathcal{B}} \underset{\sim}{y} = \underset{\sim}{\mathcal{A}} \underset{\sim}{v} ,$$

where

(3.7) $$\underset{\sim}{\mathcal{B}} = \sum_{k=0}^{p} \underset{\sim}{B}_k \otimes \underset{\sim}{L}^k , \underset{\sim}{\mathcal{A}} = \sum_{g=0}^{q} \underset{\sim}{A}_g \otimes \underset{\sim}{L}^g ,$$

and $\underset{\sim}{v}$ has the distribution $N(\underset{\sim}{0}, \underset{\sim}{\mathcal{S}})$, where

(3.8) $$\underset{\sim}{\mathcal{S}} = \underset{\sim}{\Sigma} \otimes \underset{\sim}{I}_T .$$

The logarithm of the likelihood function is

(3.9) $$\log L = -\frac{Tm}{2} \log 2\pi - \frac{T}{2} \log |\underset{\sim}{\Sigma}| - \frac{1}{2} \underset{\sim}{y}' \underset{\sim}{\mathcal{B}}' \underset{\sim}{\mathcal{A}}'^{-1} \underset{\sim}{\mathcal{S}}^{-1} \underset{\sim}{\mathcal{A}}^{-1} \underset{\sim}{\mathcal{B}} \underset{\sim}{y} .$$

The first partial derivatives of (3.9) evaluated at the initial estimates are

(3.10) $$\underset{\sim}{q}_g^{(0)} = \underset{\sim}{K}(\underset{\sim}{I}_m \otimes \underset{\sim}{L}^g \hat{\underset{\sim}{V}}_0)' \hat{\underset{\sim}{\mathcal{A}}}_0'^{-1} \hat{\underset{\sim}{\mathcal{S}}}_0^{-1} \hat{\underset{\sim}{\mathcal{A}}}_0^{-1} \hat{\underset{\sim}{\mathcal{B}}}_0 \underset{\sim}{y} ,$$

$$(3.11) \qquad \hat{p}_k^{(0)} = -K(I_m \otimes L^k Y)' \hat{\mathcal{A}}_0'^{-1} \hat{\mathcal{S}}_0^{-1} \hat{\mathcal{A}}_0^{-1} \hat{\mathcal{B}}_0 y ,$$

where

$$(3.12) \qquad \text{vec } \hat{V}_0 = \hat{v}_0 = \hat{\mathcal{A}}_0^{-1} \hat{\mathcal{B}}_0 y ,$$

$$(3.13) \qquad K = \begin{bmatrix} E_{11} & \cdots & E_{m1} \\ \vdots & & \vdots \\ E_{1m} & \cdots & E_{mm} \end{bmatrix} ,$$

and E_{ij} is the $m \times m$ matrix with 1 in the i-th row and j-th column and 0's elsewhere. (The permutation matrix K has the properties vec $A = K$ vec A', $K = K'$, and $K^2 = I_{m^2}$.) The submatrices on the left-hand side of (2.1) are

$$(3.14) \qquad \hat{\Phi}_{gh}^{(0)} = K(I_m \otimes L^g \hat{V}_0)' \hat{\mathcal{A}}_0'^{-1} \hat{\mathcal{S}}_0^{-1} \hat{\mathcal{A}}_0^{-1} (I_m \otimes L^h \hat{V}_0) K ,$$

$$(3.15) \qquad \hat{\Omega}_{g\ell}^{(0)} = -K(I_m \otimes L^g \hat{V}_0)' \hat{\mathcal{A}}_0'^{-1} \hat{\mathcal{S}}_0^{-1} \hat{\mathcal{A}}_0^{-1} (I_m \otimes L^\ell Y) K ,$$

$$(3.16) \qquad \hat{\Psi}_{k\ell}^{(0)} = K(I_m \otimes L^k Y)' \hat{\mathcal{A}}_0'^{-1} \hat{\mathcal{S}}_0^{-1} \hat{\mathcal{A}}_0^{-1} (I_m \otimes L^\ell Y) K .$$

The matrices $\hat{\Phi}_0$, $\hat{\Omega}_0$, and $\hat{\Psi}_0$ are approximately block Toeplitz. The vector $\hat{v}_0$ is not calculated by (3.12), but rather from (1.1) with initial estimates for the parameters. The first row of $\hat{V}_0$ is y_1' . The

second row is found from (1.1) for $t = 2$ (with $y_0 = \dots = y_{2-p} = 0$ and $v_0 = \dots - v_{2-q} = 0$) . Successive rows of $\hat{V}_0$ are found recursively.

Reinsel (1976) treated $y_1, \dots, y_p$ as fixed and assumed $v_{p+1-q} = \dots = v_p = 0$. The difference in the above equations is that $L^k Y$ is replaced by a $(T-p) \times m$ matrix with y'_{t-p+k} as its t-th row, L is a $(T-p) \times (T-p)$ matrix and V is a $(T-p) \times m$ matrix. In either case the asymptotic theory is the same as that given for the estimates in the frequency domain. (Reinsel did not need K because he used vec A'_g and vec B'_k .)

The nature of the procedure is that the equations for the increment in the estimates of the coefficients are analogous to weighted least squares equations for the regression of y on $L^g V_0$ and $L^\ell Y$.

Osborn (1977) has shown how the exact likelihood can be expressed for the pure moving average process and shows how it can be evaluated for values of $A_1, \dots, A_q$ and Σ .

4. Initial estimates. We state initial estimates in terms of the observed covariances $\{c_h\}$ given by (1.9) . In practice the mean of the process is unknown and in (1.9) y_t (and y_{t+h}) would be replaced by $y_t - \bar{y}$ (and $y_{t+h} - \bar{y}$, respectively), where $\bar{y} = (1/T)\Sigma_{t=1}^T y_t$. Then initial estimates of $B_1, \dots, B_p$ are obtained from

$$\sum_{k=1}^{p} \hat{B}_k^{(0)} c_{k-\ell} = - c_{-\ell} , \qquad \ell = q+1, \dots, q+p . \tag{4.1}$$

Then we can form (with $y_0 = \dots = y_{1-p} = 0$ and $\hat{B}_0^{(0)} = I$)

$$\hat{u}_t^{(0)} = \sum_{k=0}^{p} \hat{B}_k^{(0)} y_{t-k} , \qquad t = 1, \dots, T , \tag{4.2}$$

$$(4.3)\qquad \underset{\sim}{c}^{u}_{h0} = \frac{1}{T}\sum_{t=1}^{T-h} \hat{\underset{\sim}{u}}^{(0)}_{t}\ \hat{\underset{\sim}{u}}^{(0)\prime}_{t+h} = - \underset{\sim}{c}^{u\prime}_{h0}, \quad h = 0, 1, \ldots, T-1,$$

$$(4.4)\qquad \hat{\underset{\sim}{f}}^{u}_{0}(\lambda) = \frac{1}{2\pi}\sum_{h=-q}^{q} e^{i\lambda h}\, \underset{\sim}{c}^{u}_{h0}\ .$$

If $\hat{\underset{\sim}{f}}^{u}_{0}(\lambda)$ is a Hermitian nonnegative definite matrix function, it can be factored according to (2.9) to define $\hat{\underset{\sim}{A}}^{(0)}_{1}, \ldots, \hat{\underset{\sim}{A}}^{(0)}_{q}$ and $\hat{\underset{\sim}{\Sigma}}_{0}$ [Robinson (1967)]. Alternatively, $\underset{\sim}{A}^{(0)}_{1}, \ldots, \underset{\sim}{A}^{(0)}_{q}$ can be found from

$$(4.5)\qquad \sum_{h=1}^{q}\sum_{t=1}^{T} \hat{\underset{\sim}{f}}^{u}_{0}(\lambda_t)^{-1}\, \underset{\sim}{I}^{u}_{0}(\lambda_t)\hat{\underset{\sim}{f}}^{u}_{0}(\lambda_t)^{-1}\, \hat{\underset{\sim}{A}}^{(0)}_{h}\, e^{i\lambda_t(h-g)}$$

$$= - \sum_{t=1}^{T} \hat{\underset{\sim}{f}}^{u}_{0}(\lambda_t)^{-1}\, \underset{\sim}{I}^{u}_{0}(\lambda_t)\, \hat{\underset{\sim}{f}}^{u}_{0}(\lambda_t)^{-1}\, e^{-i\lambda_t g}, \quad h = 1, \ldots, q,$$

where

$$(4.6)\qquad \underset{\sim}{I}^{u}_{0}(\lambda) = \frac{1}{2\pi}\sum_{h=-(T-1)}^{T-1} e^{i\lambda h}\, c^{u}_{h0}\ ,$$

as suggested by Hannan. Then

$$(4.7)\qquad \hat{\Sigma}_{0} = \frac{2\pi}{T}\sum_{t=1}^{T} \hat{\underset{\sim}{A}}_{0}(e^{i\lambda_t})^{-1}\, \underset{\sim}{I}^{u}_{0}(\lambda_t)\hat{\underset{\sim}{A}}^{*}_{0}(e^{i\lambda_t})^{-1}\ .$$

REFERENCES

Akaike, Hirotugu (1973), "Maximum Likelihood Identification of Gaussian Autoregressive Moving Average Models," Biometrika, Vol. 60, pp. 255-265.

Anderson, T. W. (1971), The Statistical Analysis of Time Series, John Wiley & Sons, Inc., New York.

Anderson, T. W. (1977), "Estimation for Autoregressive Moving Average Models in the Time and Frequency Domains," Annals of Statistics, Vol. 5, pp. 842-865.

Dunsmuir, W., and Hannan, E. J. (1976), "Vector Linear Time Series Models," Advances in Applied Probability, Vol. 8, pp. 339-364.

Hannan, E. J. (1969a), "The Identification of Vector Mixed Autoregressive-Moving Average Systems," Biometrika, Vol. 56, pp. 223-225.

Hannan, E. J. (1969b), "The Estimation of Mixed Moving Average Autoregressive Systems," Biometrika, Vol. 56, pp. 579-593.

Hannan, E. J. (1970), Multiple Time Series, John Wiley & Sons, Inc., New York.

Hannan, E. J. (1971), "The Identification Problem for Multiple Equation Systems with Moving Average Errors," Econometrica, Vol. 39, pp. 751-765.

Kashyap, R. L., and Nasburg, Robert E. (1974), "Parameter Estimation in Multivariate Stochastic Difference Equations," Transactions on Automatic Control, Vol. AC-19, pp. 784-797.

Marcus, Marvin, and Minc, Henryk (1964), A Survey of Matrix Theory and Matrix Inequalities, Prindle, Weber, and Schmidt, Inc., Boston.

Nicholls, D. F. (1976), "The Efficient Estimation of Vector Linear Time Series Models," Biometrika, Vol. 63, pp. 381-390.

Nicholls, D. F. (1977), "A Comparison of Estimation Methods for Vector Linear Time Series Models," Biometrika, Vol. 64, pp. 85-90.

Osborn, Denise R. (1977), "Exact and Approximate Maximum Likelihood Estimators for Vector Moving Average Processes," Journal of the Royal Statistical Society, Series B, Vol. 39, pp. 114-118.

Reinsel, Gregory C. (1976), "Maximum Likelihood Estimation of Vector Autoregressive Moving Average Models," Technical Report No. 117, Department of Statistics, Carnegie-Mellon University, Pittsburgh.

Robinson, E. A. (1967), Multichannel Time Series Analysis with Digital Computer Programs, Holden-Day, San Francisco.

Tunnicliffe Wilson, G. (1973), "The Estimation of Parameters in Multivariate Time Series Models," Journal of the Royal Statistical Society, Series B, Vol. 35, pp. 76-85.

TECHNICAL REPORTS

OFFICE OF NAVAL RESEARCH CONTRACT N00014-67-A-0112-0030 (NR-042-034)

1. "Confidence Limits for the Expected Value of an Arbitrary Bounded Random Variable with a Continuous Distribution Function," T. W. Anderson, October 1, 1969.

2. "Efficient Estimation of Regression Coefficients in Time Series," T. W. Anderson, October 1, 1970.

3. "Determining the Appropriate Sample Size for Confidence Limits for a Proportion," T. W. Anderson and H. Burstein, October 15, 1970.

4. "Some General Results on Time-Ordered Classification," D. V. Hinkley, July 30, 1971.

5. "Tests for Randomness of Directions against Equatorial and Bimodal Alternatives," T. W. Anderson and M. A. Stephens, August 30, 1971.

6. "Estimation of Covariance Matrices with Linear Structure and Moving Average Processes of Finite Order," T. W. Anderson, October 29, 1971.

7. "The Stationarity of an Estimated Autoregressive Process," T. W. Anderson, November 15, 1971.

8. "On the Inverse of Some Covariance Matrices of Toeplitz Type," Raul Pedro Mentz, July 12, 1972.

9. "An Asymptotic Expansion of the Distribution of "Studentized" Classification Statistics," T. W. Anderson, September 10, 1972.

10. "Asymptotic Evaluation of the Probabilities of Misclassification by Linear Discriminant Functions," T. W. Anderson, September 28, 1972.

11. "Population Mixing Models and Clustering Algorithms," Stanley L. Sclove, February 1, 1973.

12. "Asymptotic Properties and Computation of Maximum Likelihood Estimates in the Mixed Model of the Analysis of Variance," John James Miller, November 21, 1973.

13. "Maximum Likelihood Estimation in the Birth-and-Death Process," Niels Keiding, November 28, 1973.

14. "Random Orthogonal Set Functions and Stochastic Models for the Gravity Potential of the Earth," Steffen L. Lauritzen, December 27, 1973.

15. "Maximum Likelihood Estimation of Parameters of an Autoregressive Process with Moving Average Residuals and Other Covariance Matrices with Linear Structure," T. W. Anderson, December, 1973.

16. "Note on a Case-Study in Box-Jenkins Seasonal Forecasting of Time series," Steffen L. Lauritzen, April, 1974.

17. "General Exponential Models for Discrete Observations," Steffen L. Lauritzen, May, 1974.

18. "On the Interrelationships among Sufficiency, Total Sufficiency and Some Related Concepts," Steffen L. Lauritzen, June, 1974.

19. "Statistical Inference for Multiply Truncated Power Series Distributions," T. Cacoullos, September 30, 1974.

Office of Naval Research Contract N00014-75-C-0442 (NR-042-034)

20. "Estimation by Maximum Likelihood in Autoregressive Moving Average Models in the Time and Frequency Domains," T. W. Anderson, June 1975.

21. "Asymptotic Properties of Some Estimators in Moving Average Models," Raul Pedro Mentz, September 8, 1975.

22. "On a Spectral Estimate Obtained by an Autoregressive Model Fitting," Mituaki Huzii, February 1976.

23. "Estimating Means when Some Observations are Classified by Linear Discriminant Function," Chien-Pai Han, April 1976.

24. "Panels and Time Series Analysis: Markov Chains and Autoregressive Processes," T. W. Anderson, July 1976.

25. "Repeated Measurements on Autoregressive Processes," T. W. Anderson, September 1976.

26. "The Recurrence Classification of Risk and Storage Processes," J. Michael Harrison and Sidney I. Resnick, September 1976.

27. "The Generalized Variance of a Stationary Autoregressive Process," T. W. Anderson and Raul P.Mentz, October 1976.

28. "Estimation of the Parameters of Finite Location and Scale Mixtures," Javad Behboodian, October 1976.

29. "Identification of Parameters by the Distribution of a Maximum Random Variable," T. W. Anderson and S.G. Ghurye, November 1976.

30. "Discrimination Between Stationary Guassian Processes, Large Sample Results," Will Gersch, January 1977.

31. "Principal Components in the Nonnormal Case: The Test for Sphericity," Christine M. Waternaux, October 1977.

32. "Nonnegative Definiteness of the Estimated Dispersion Matrix in a Multivariate Linear Model," F. Pukelsheim and George P.H. Styan, May 1978.

TECHNICAL REPORTS (continued)

33. "Canonical Correlations with Respect to a Complex Structure," Steen A. Andersson, July 1978.

34. "An Extremal Problem for Positive Definite Matrices," T.W. Anderson and I. Olkin, July 1978.

35. " Maximum likelihood Estimation for Vector Autoregressive Moving Average Models," T. W. Anderson, July 1978.

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER 35	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) MAXIMUM LIKELIHOOD ESTIMATION FOR VECTOR AUTOREGRESSIVE MOVING AVERAGE MODELS		5. TYPE OF REPORT & PERIOD COVERED Technical Report
		6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) T. W. ANDERSON		8. CONTRACT OR GRANT NUMBER(s) N00014-75-C-0442
9. PERFORMING ORGANIZATION NAME AND ADDRESS Department of Statistics Stanford University Stanford, California		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS (NR-042-034)
11. CONTROLLING OFFICE NAME AND ADDRESS Office of Naval Research Statistics & Probability Program Code 436 Arlington, Virginia 22217		12. REPORT DATE July 1978
		13. NUMBER OF PAGES 16
14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office)		15. SECURITY CLASS. (of this report) UNCLASSIFIED
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

APPROVED FOR PUBLIC RELEASE; DISTRIBUTION UNLIMITED.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Maximum likelihood estimation, vector autoregressive moving average, Newton-Raphson, scoring, information matrix, time series analysis.

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

SEE REVERSE SIDE

20. ABSTRACT

The vector autoregressive moving average model is a multivariate stationary stochastic process $\{\underset{\sim}{y}_t\}$ satisfying

$$\sum_{k=0}^{p} \underset{\sim}{B}_k \underset{\sim}{y}_{t-k} = \sum_{g=0}^{q} \underset{\sim}{A}_g \underset{\sim}{v}_{t-g} ,$$

where the unobservable multivariate process $\{\underset{\sim}{v}_t\}$ consists of independently identically distributed random vectors. The coefficient matrices and the covariance matrix of $\underset{\sim}{v}_t$ are to be estimated from an observed sequence $\underset{\sim}{y}_1, \ldots, \underset{\sim}{y}_T$. Under the assumption of normality the method of maximum likelihood is applied to likelihoods suitably modified for techniques in the frequency and time domains. Newton-Raphson and scoring iterative methods are presented.